DOCUMENT RESUME

ED 215 330 CS 006 641

AUTHOR Duffy, Thomas M.; U' Ren, Paula Kabance
TITLE The Role of Word Difficulty and Sentence Length in Text Comprehension.
INSTITUTION Navy Personnel Research and Development Center, San Diego, Calif.
REPORT NO NPRDC-TR-82-44
PUB DATE May 82
NOTE 36p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Aptitude; Adults; *Language Processing; Military Personnel; *Readability; *Reading Comprehension; *Reading Difficulties; *Reading Research; *Sentence Structure; Textbook Preparation
IDENTIFIERS *Language Simplification

ABSTRACT

The relationship of readability, "readable writing" techniques, and comprehension was evaluated in a series of five experiments involving United States Navy recruits. Eight expository passages from a standardized reading test were revised by using word lists to simplify the vocabulary and a restriction in syntactic structure to simplify the sentences. The passages were presented to recruits in five different conditions in which the reading task (reading-to-do or reading-to-learn), the time allowed for reading, and the comprehension test format (multiple choice or cloze) were varied. Results showed that, with the exception of one instance, the readable writing revisions had no practical effect on comprehension regardless of the reading skills of the recruits. The one positive instance of practical significance was for low ability readers when the reading task was in a reading-to-learn format. The results indicated that readable writing revisions can facilitate comprehension only under very particular circumstances, and that changes in the readability score are not in any way predictive of the changes in comprehension. Thus, a readability formula score is neither an effective guideline nor an effective criterion to impose on a writer. (Materials used in the experiments are appended.) (FL)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

NPRDC TR 82-44

May 1982

ED215330

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

# THE ROLE OF WORD DIFFICULTY AND SENTENCE LENGTH IN TEXT COMPREHENSION

Thomas M. Duffy
Paula Kabance U' Ren

Reviewed by
E. G. Aiken

Released by
James F. Kelly, Jr.
Commanding Officer

Navy Personnel Research and Development Center
San Diego, California 92152

C S806 641

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>NPRDC TR 82-44 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>THE ROLE OF WORD DIFFICULTY AND SENTENCE LENGTH IN TEXT COMPREHENSION | | 5. TYPE OF REPORT & PERIOD COVERED<br>Technical Report<br>FY 77-81<br>6. PERFORMING ORG. REPORT NUMBER<br>15-81-20 |
| 7. AUTHOR(s)<br>Thomas M. Duffy<br>Paula Kabance U' Ren | | 8. CONTRACT OR GRANT NUMBER(s) |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Navy Personnel Research and Development Center<br>San Diego, California 92152 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>ZF63-522-001-011-03.01 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Navy Personnel Research and Development Center<br>San Diego, California 92152 | | 12. REPORT DATE<br>May 1982<br>13. NUMBER OF PAGES<br>35 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>UNCLASSIFIED<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Readability
Comprehension
Technical manuals
Technical documentation
Writing
Editing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The relationship of readability, readable writing techniques, and comprehension was evaluated in a series of five experiments. Passages from a standard reading test were rewritten by following specific basic readable writing rules for simplifying words and sentences. Following the procedures resulted in major improvements in readability (six grade levels). However, except for one instance, this manipulation had no practical effect on comprehension across a series of five experiments in which the skill of the readers, the

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

time allowed for reading, and the type of comprehension test were varied. Comprehension was facilitated only when the vocabulary alone was simplified and then only for low ability readers taking a reading-to-learn type of test. The results indicate the ineffectiveness of readable writing rules and readability formulas for regulating or monitoring the comprehension difficulty of text. Alternative approaches are discussed.

## FOREWORD

This research was performed under exploratory development task area ZF63-522-001-011 (The Assessment and Enhancement of Prerequisite Skills), work unit number 03.01 (Language Skills: Assessment and Enhancement). The report is one of several examining reading requirements, reading skill levels, and the effects of a mismatch between reading skill and requirements on school and job performance in the Navy (e.g., NPRDC SR 78-19 and NPRDC TR 77-40). It evaluates and finds deficiencies in the use of readability techniques as a means of improving the comprehensibility of text.

The findings should be of particular interest to those involved in the production or acquisition of training, job, or general information text materials and manuals.

JAMES F. KELLY, JR.
Commanding Officer

JAMES J. REGAN
Technical Director

## SUMMARY

### Problem

With the rapid growth in the volume and the importance of technical documentation, it is becoming increasingly important that technical and training manuals be easy to use. An important factor in determining ease of use, which has been especially difficult to effectively describe or control, has been the comprehensibility of the materials.

### Background

All of the services have turned increasingly to the use of readability formulas to guide and regulate comprehensibility of prose in the production of technical documents. The readability formula is an empirically-based method for predicting comprehension of a text from a count of certain of its word and sentence characteristics. These formulas are used as guidelines or feedback to the writer, identifying the expected comprehension difficulty of a draft, and as criteria that the writer must achieve. Writers are not expected to achieve the readability criterion simply by writing to the formula (e.g., by using a word with fewer syllables regardless of whether or not it is a "better" word). Rather, they are expected to follow readable writing guidelines (e.g., use more familiar words and direct simple sentences). This application of the readability formula presumes a strong causal relationship between readability and comprehension and readable writing techniques and comprehension. However, there is little support for these causal assumptions.

### Purpose

The purpose of this research was to determine the extent to which increasing readability by carrying out readable writing revisions will improve comprehension.

### Method

Eight passages from a reading test were revised by using word lists to simplify the vocabulary and a restriction in syntactic structure to simplify the sentences. Vocabulary simplification reduced the readability formula score from a reading grade level (RGL) of 11.3 to 9.6, while sentence simplification resulted in an average 7.7 RGL for the passages. When both words and sentences were simplified, the readability was reduced to 5.5 RGL--a reduction of six RGLs from the original materials.

These passages were presented to Navy recruits in a series of five experiments in which the reading task (reading-to-do vs. reading-to-learn), the time allowed for reading, and the comprehension tests (multiple-choice vs. cloze) were varied.

### Findings

Except for one instance, the "readable writing" revisions had no practical effect on comprehension regardless of the reading skills of the recruits. The one positive instance of practical significance was for low ability readers when the reading task was in a reading-to-learn format (i.e., the reader did not know what questions would be asked until the material had been read and removed). Even here, however, only vocabulary simplifications--the text manipulation that resulted in the smallest change in the readability score--produced a significant effect.

## Conclusions

The results indicate that "readable writing" revisions can facilitate comprehension under very particular circumstances. However, the resulting changes in the readability score are not in any way predictive of the changes in comprehension. Thus, a formula score is neither an effective guideline nor an effective criterion to impose on a writer.

## Recommendation

The text production process should be studied to determine how both government and contractor management of the production effort can be modified to increase attention to comprehensibility. This should include a consideration of computer-aided authoring, management priorities, and personnel qualifications. The results of this study fail to support the use of a readability formula as a means of controlling the comprehensibility of text.

## CONTENTS


| | Page |
|---|---|
| INTRODUCTION | 1 |
| Background and Problem | 1 |
| Purpose | 4 |
| GENERAL METHOD | 4 |
| Design and Materials | 4 |
| Subjects and Testing Procedures | 6 |
| EXPERIMENT 1 | 6 |
| Method | 6 |
| Results | 7 |
| EXPERIMENT 2 | 8 |
| Method | 9 |
| Results and Discussion | 9 |
| EXPERIMENT 3 | 11 |
| Method | 11 |
| Results | 12 |
| EXPERIMENT 4 | 13 |
| Method | 13 |
| Results and Discussion | 14 |
| EXPERIMENT 5 | 15 |
| Method | 15 |
| Results and Discussion | 15 |
| DISCUSSION AND CONCLUSIONS | 17 |
| RECOMMENDATION | 20 |
| REFERENCES | 21 |
| APPENDIX--ORGINAL AND THREE REVISIONS OF ONE OF THE EXPERIMENTAL PASSAGES FROM THE NELSON-DENNY TEST | A-0 |
| DISTRIBUTION LIST | |

## LIST OF FIGURES


Page

1. Percent correct under each revision condition, Experiment 1. . . . . 8

2. Percent correct under each revision condition for medium and low ability groups, Experiment 2. . . . . 10

3. Proportion correct (adjusted means) under exact cloze scoring for each revision condition, Experiment 3. . . . . 12

4. Percent correct under each revision condition for high and low ability groups, Experiment 4. . . . . 14

# INTRODUCTION

## Background and Problem

Written text, whether it is in the traditional book form or as microfilm or computer display, plays a central role throughout society as the primary means of communicating and documenting information. The rapid expansion of technology with its associated knowledge requirements has greatly increased both the volume and the importance of text documentation. All facets of society have entered the information age and, as a consequence, it is seldom the case today that "experts" in any field of endeavor can possess all the necessary knowledge "in their head." Instructional textbooks are growing in number and in degree of specialization. Professional journals are dividing and multiplying.

The incredible growth in the requirements for text documentation and the effects on the individual can perhaps best be illustrated by the military's technical manual requirements. According to estimates, the U.S. Navy alone has 25 million pages of technical manuals (not including training texts) and 400,000 pages are added or revised yearly (Sulit & Fuller, 1976). In the last 25 years, the number of pages required to document a modern weapon system has increased by 14,000 percent (Muller, 1976). This growth in materials obviously has not been accompanied by a comparable growth in the number of naval personnel. In fact, manning levels in some ship classes, including destroyers, have decreased (Aiken, 1980). Thus, there is a tremendous increase in the documentation that any one individual must utilize. When this increasing requirement for information is added to the national decline in literacy skills of the users and the high employee turnover in the military as well as in industry, it is obvious that information texts must be designed for maximum usability.

The usability of a text can be described in terms of four dimensions: access, accuracy, completeness, and comprehensibility. The reader must be able to locate the relevant information. Once located, the information must be accurate and be described to a level of detail that matches the information requirement and the knowledge level of the reader. However, even if the information is complete, it may not be usable if it is written in a convoluted, jargon-ridden fashion and is not supported by adequate graphics. While all four dimensions are important to usability, the last dimension--comprehensibility--has been the most troublesome and is currently receiving the greatest attention.

There has been a nationwide cry for clear, comprehensible writing. The focus of this concern has been on consumer documents and on industrial and military manuals. These types of documents pose a special problem because they are not sold as separate items but, rather, are given away or accompany some other, more major purchase. Hence, they are not subject to the typical marketplace controls for product quality. The legal document that comes with the insurance policy the agent described is simply accepted. Similarly, if customers want a particular piece of equipment--be it a modern toy, a mixer, or a radar set, they must accept the accompanying assembly, use, and maintenance instructions. If they want the primary product, they simply accept the documentation. From the producer's point of view, the customer is purchasing the primary product; thus, the features and the cost of the product are most important. There is little incentive to improve the quality of the documentation; indeed, there is a disincentive since such efforts would only add to the cost of the primary product.

It is obvious that, under these conditions, there is little reason for the producers of these types of documents to expend the time or the money to ensure adequate comprehensibility. Although it can be demonstrate time and again that documents can be

redesigned to make them more comprehensible (see any issue of Simply Stated, Battison & Landesman, 1981), the difficulty is to get the publishers to produce that type of document under normal conditions and funding. If this is to be achieved with those producing documents for the Navy, then either some means of regulation or some type of incentive system is required. An incentive system is unlikely since this will simply add to the cost. Regulations require some sort of assessment system and, unfortunately, comprehensible writing is not readily described, prescribed, or assessed by objective rules and procedures. Testing of personnel to determine comprehensibility is typically too costly and not feasible logistically. Consider, for example, the requirements for testing personnel on the 400,000 pages of technical manual text produced annually by the Navy alone. The use of judges to determine the comprehensibility of text would appear to be a reasonable alternative to direct testing. However, judges simply do not agree on their rankings of comprehensibility (Carver, 1974; Wright[1]). Further, even experienced writers disagree and contradict each other in their rules for comprehensible writing (Klare, undated). Thus, it would seem that comprehensible writing currently is largely an art not readily subject to objective description.

Despite these difficulties, the requirement for a legally binding comprehensibility criterion remains. To meet this requirement, many organizations have turned to the readability formula as a text production guideline and criterion. Pressman (1979) reports that two thirds of the states requiring clear writing in consumer regulations have specified a readability formula score as the criterion to be met. Similarly, military specifications for the preparation of manuals are increasingly invoking a readability formula score to "match the manual to the people" (Department of the Air Force, 1977; Department of the Army, 1978; Department of Defense, 1978, 1980). While the readability formula is an objective device, it is not based on the specifications of particular writing rules. Furthermore, formulas are inexpensive and easy to apply. Thus, it would seem that their application has potential as a primary technique for regulating the production of comprehensible text. The readability formula, of course, only attends to the writing style of the prose. It does not address graphics, graphic-prose coordination, format, or organization. Nonetheless, if a readability formula is effective as a criterion for the prose, then at least a significant portion of the battle to obtain comprehensible text will have been won.

While readability formulas are objective and easy to apply during text production, it is still not clear if their use as a criterion or as feedback in the production process is valid. These formulas have all been developed as empirical tools to predict comprehension of text that is already written. No model or theory of comprehension underlies the text variables used to predict comprehension. They are simply the best predictors that may or may not be causally associated with comprehension. Yet, obviously, there is an implicit causal assumption underlying the application of a formula in regulating the production of text.

Klare (1979) has argued that readability formulas can be effective aids in preparing more comprehensible text, as long as the writer does not write to the formula. That is, Klare recognizes that the readability formula is a predictive tool and simply changing text to meet the requirements of a formula (i.e., by shortening words and sentences) will not necessarily result in more comprehensible text, even if a lower readability score is achieved. Writing to the formula would only work if one presumed that the particular predictor variables (e.g., number of letters per word) were also the causal variables.

---

[1]Wright, P., personal communication, November 1980.

While Klare (1979) does not advocate writing to the formula, he does advocate the use of "readable writing" techniques both to reduce the formula score and increase comprehensibility. A readable writing approach, while not necessarily addressing the formula variable, is focused exclusively on variables at the word and sentence level (Klare, 1975). Thus, instead of mechanically using shorter words, more familiar and concrete words, which, in general, will be shorter, are sought. Thus, both a readable writing approach and writing to formula result in shorter words and sentences. They are perhaps best distinguished by the attitude of the writer--does he/she search for a "good" short word or simply a short word?

A readable writing approach is, in fact, the only option (outside of writing to the formula) open to a writer to revise the draft of a text to meet a readability criterion. Since the predictor variables are virtually always at the word and sentence level, revision must address word and sentence factors. Improving the format, reorganizing, adding examples, and defining objectives are all revision techniques that might improve comprehension, but are irrelevant to a readable writing approach and would have little effect on a readability score. Thus, the use of readability feedback or a readability criterion, along with readable writing techniques, presumes that the primary causal factors in comprehension are at the word and sentence level. The fact that readability predictors correlate so strongly with comprehension of already prepared text might lead one to accept such a presumption. However, it may be that a good writer is a conscientious writer; he/she not only uses simple words and sentences but also organizes well, presents examples and transition information, etc. If so, then the use of a readability formula criterion will not lead to significant improvements in comprehensibility even if readable writing techniques are used.

There is a wealth of data (e.g., Adams, 1967) indicating the importance of word and sentence variables in learning and recall. Because of these data, one might presume that readable writing techniques would be effective, if not wholly as effective as the readability formula multiple regression values would suggest. However, as the data are primarily from verbal learning studies, surprisingly few are from large text segments. Goetz (1975), in a recent review, reports an "embarrassing" lack of empirical evidence on learning and recall differences between sentence lists and text. Klare (1963) reports only six readability studies involving the controlled manipulation of word or sentence difficulty. Only one of those six studies found a positive effect and this was for simplifying vocabulary. In other studies, Tuinman and Brady (1973), rather than substituting more familiar words to simplify text, made unfamiliar vocabulary in the text more comprehensible to the readers by giving extensive vocabulary training sessions. While the training led to a 20-percent improvement in vocabulary knowledge, there was no effect on text comprehension. Coleman (1962), using a section of a psychology textbook, increased sentence length to an average of 38.6 words in one condition and shortened sentences to an average of 15.4 words in another condition. Although he used Flesch's (1948) guidelines for readable writing to revise the passages, this very large sentence length manipulation resulted in only marginally significant differences in comprehension.

Siegel, Lautman, and Burkett (1974) simplified both the vocabulary and syntax of technical materials following readable writing guidelines. While the simplification resulted in a 3.5 RGL improvement in readability, there was a marginally significant effect on comprehension in only one of three experiments. Kniffen, Stevenson, Klare, Entin, Slaughter, and Hooke (1979) also failed to find any effect on comprehension when they simplified technical materials, even when readability was four grade levels less than the RGL of the participants.

Klare (1976) presents a model of the text use situation that places limits on the circumstances under which readability improvement will result in higher comprehension.

Klare proposes that manipulation in readability may be ineffective if the reader is highly motivated, familiar with the subject area, or has excessive reading time. In addition, he suggests that improved comprehension requires the manipulation of both the sentence and the word difficulty variables as well as large differences in readability that bracket the skill level of the reader. From his review of readability research, Klare (1976) concludes that most null results in readability manipulation can be accounted for by one of these factors.

Purpose

The present research sought to test the effectiveness of readability revisions as a means of producing more comprehensible text. Revisions for readability were not written to formula directly but, rather, by using "readable writing" guidelines. The objective was to achieve large changes in readability of the passages by following strict readable writing guidelines, thus simulating what a writer would try to do in revising for readability. In addition, the purpose was to test the effect of the readability manipulations under the most favorable conditions possible, thus maximizing the potential for a "readability effect." In line with Klare's (1976) model, readers were unfamiliar with the topics and had low motivation. Further, their reading skill varied over a wide range so that the interaction of reading skill and readability could be assessed.

## GENERAL METHOD

Design and Materials

Eight passages from the Nelson-Denny Reading Test Form B, Part II (Reading Comprehension) (Nelson & Denny, 1960) provided the basic materials for all of the experiments. Four narrative and four expository passages are in the test. The first passage consists of 227 words, while all of the others are approximately 200 words long. The Nelson-Denny test questions--all five-choice multiple choice--were also used in most of the experiments. There were eight questions associated with the long passage; and four, with each of the others.

Five experiments were conducted. In each, there were three factors--one subject-related and two passage-based. The vocabulary and the sentences of each passage were manipulated independently to create the two independent passage factors--sentence complexity and vocabulary difficulty. There were two levels for each factor, original and simplified, thus yielding four versions of each passage. The various versions for one of the passages is presented in the appendix. Ability, the subject-related factor in each experiment, was assessed by performance on an alternate form (Form A) of the Nelson-Denny Reading Test (Nelson & Denny, 1960).

For the vocabulary simplification condition (Condition V), the words in all Nelson-Denny passages and in the multiple-choice questions had to meet one of the following criteria: (1) A or AA on the Thorndike-Lorge word list (Thorndike & Lorge, 1944), (2) no higher than the fourth grade level on the Durell (1956) word list, and (3) included in the Dale list of 400 words (Dale & Chall, 1948). Twenty-six percent of the content words failed to meet these criteria and were changed by substituting words from one of the lists. This amounted to changing 13 percent of the total number of words in the passages. Excluding proper nouns, there were only eight words that failed to meet the criteria and for which contextually appropriate substitutes could not be found. These words were unchanged. Substituting more frequent and easier words also resulted in shortening the average word length. The mean number of syllables per word was 3.18 for words removed and 1.84 for words added, a 42-percent reduction in syllable length for changed words.

For sentence simplification (Condition S), the criterion that every sentence be a simple sentence with no more than two phrases was applied to the passages and multiple-choice questions. Pronouns, articles, and connectives were added and deleted as necessary to meet the criterion and to maintain the continuity and flow of the stories. Applying this syntactic criterion resulted in shortening the average sentence length from 21.3 to 9.8 words, a reduction of more than 50 percent.

Finally, both the vocabulary and the sentence simplification criteria were applied to the passages and questions (Condition S-V). It is this joint manipulation of the readability factors that Klare (1976) considered essential for producing comprehension effects.

The effects of these manipulations on readability scores were determined by applying three readability formulas to each passage in each of the experimental conditions as well as to the passages in their original form (Condition O). The formulas used were the Kincaid revision of the Flesch formula for use with the military (Kincaid, Fishburne, Rogers, & Chissom, 1975), the FOG count (Gunning, 1952), and the Fry formula (Fry, 1968). As can be seen in Table 1, all three formulas clearly indicate that the simplifications significantly increased readability with the effect being greater for the sentence (Condition S) than for the vocabulary (Condition V) manipulation. The distribution of Kincaid-Flesch scores between Conditions O and S-V do not overlap. Thus, the manipulations were effective in producing large differences in readability as required if comprehension effects are to be obtained.

Table 1

Changes in Readability as a Function of Word and Sentence Modifications

| Revision Condition | Words Per Sentence (Mean) | Percent of Content Words Changed | Reading Grade Levels (RGLs) (Mean and Range) | | |
|---|---|---|---|---|---|
| | | | Kincaid-Flesch | FOG Count | Fry[a] |
| Original (O) | 21.3 | 0 | 11.5<br>8.5-16.6 | 15.1<br>8.0-21.4 | 11.1<br>7.0-College |
| Vocabulary (V) | 22.0 | 26 | 10.1<br>6.7-14.8 | 12.8<br>7.9-17.8 | 8.5<br>6.0-12.0 |
| Sentence (S) | 9.8 | 0 | 7.3<br>4.2-11.2 | 9.9<br>6.6-13.3 | 9.3<br>4.0-College |
| S-V | 10.2 | 26 | 5.5<br>3.4-8.2 | 7.4<br>5.3-9.6 | 6.0<br>3.0-9.0 |

[a]Fry scores of "college level" were counted as 14.0 RGL for averaging.

Only four of the multiple-choice questions required minor syntactic simplification to meet the criterion. Thus, the syntax of the questions was virtually the same for all conditions. However, significant changes in vocabulary were required to meet the vocabulary criterion. The terminology used in the passages and in the questions was kept consistent. Therefore, the vocabulary of the questions for Conditions V and S-V was different from that under Conditions O and S.

Subjects and Testing Procedures

The participants in all the experiments were male Navy recruits approximately halfway through training at the Recruit Training Command, San Diego. Recruits at San Diego are typically between 17 and 20 years old and approximately 80 percent have at least a high school diploma.

Participants were tested in groups of 40 to 70 in a large test hall having individual carrels. Each session was monitored by three experimenters. A session began with pretesting for ability using Part II (Comprehension) of Form A of the Nelson-Denny test (i.e., an alternate form of the basic experimental materials). The test was administered using the published test administration procedures (Nelson & Denny, 1960). The experimental materials were then distributed with all experimental conditions equally represented.

## EXPERIMENT 1

Sticht, Fox, Hauke, and Zapf (1977) have distinquished between two types of reading tasks: reading-to-do and reading-to-learn. A reading-to-do task involves reading with a specific, narrowly defined objective for immediate application; reading is a subtask of a specific job task. In the reading-to-learn task, the objective is to store and retain what is read for use at another time.

In this experiment, the effects of the readability manipulation were evaluated on reading-to-do tasks. Sticht et al. (1977), in interviewing Navy personnel, found that 70 percent of the reading tasks of job incumbents were reading-to-do. Of course, if one considers the use of technical manuals, the tasks are almost exclusively reading-to-do. Thus, if readability formulas are to be effective production tools for the military, readable writing techniques must improve comprehension in reading-to-do tasks.

In the reading-to-do task for Experiment 1, the participants were given the opportunity to study the four or eight questions on the passages before reading the passage. After reading was completed, the questions were presented again for the participants to answer. Thus, memory requirements were minimized and the reader knew the specific information objective before reading the passage.

Method

A total of 230 recruits were tested, with 54 to 60 assigned to each of the four conditions. After pretesting, each participant received one of the experimental sets of

booklets. In each booklet, materials for each passage were sequenced on successive pages as follows: the multiple-choice questions, a blank page, the passage, a blank page, and the multiple-choice questions. The order of the questions and the alternatives within a question was randomized from first to second presentation.

Participants were told that this was a reading test like the pretest and that the type of passages and questions would be similar. The differences in the control of reading time (per passage rather than a fixed time for the entire test) and the sequencing of questions and passages were explained. During the test, the participants turned to successive pages on cue from the experimenter. Ten seconds per question were allowed for previewing. Thus, for example, when there were four questions on a passage, the question page was presented for 40 seconds. After reading, the question page was presented for 15 seconds per question. The extra time was to allow for marking the answer on a machine-scorable answer sheet. Passages were presented to allow for a single reading at a rate of 100 words per minute.

Results

Pretest scores were used as the ability measure. Participants were divided into high and low ability groups based on the overall pretest median score of 15.5 (an RGL of 10.3 on the test norms). The sample size in the cells of the resulting 2x2x2 matrix ranged from 17 to 37. The pretest scores were analyzed using an unweighted means analysis of variance (ANOVA) to determine if random assignment of participants to the various passage revision conditions resulted in groups of equivalent reading skills.

The only source of variance to reach significance was the main effect of ability ($F = 39.0$, $df = 1,222$, $p < .001$), simply reflecting the intentional high/low ability grou[illegible] Thus, the ability levels across each text condition were equal prior to experim[illegible] treatment.

The total number of correct answers to the questions on the experimental passages was evaluated using a 2x2x2 unweighted means ANOVA. The only significant effects were the main effects due to ability ($F (1,122) = 101.8$, $p < .001$) and vocabulary ($F (1,222) = 5.2$, $p < .02$). The main effect of ability, which was expected, simply indicated that high ability readers comprehend more than do low ability readers. The vocabula[illegible] effect, while statistically significant, was of only minor practical significance. The mean percent correct without vocabulary simplification was 60.0. Simplifying vocabulary only raised performance by three percentage points--to 63.3 percent.

The results, summarized across ability levels, are presented in Figure 1. Most surprising is the lack of any significant interactions. In particular, an interaction of the readability variables with reader ability was expected. Since the high ability readers all scored above 10.3 RGL on the Nelson-Denny test, they were reasonably good comprehenders. As Klare (1976) reports, readability manipulations are not as effective for "good" readers. However, since the low ability readers had a median pretest RGL score of 8.7, these readability manipulations would surely affect their performance. Yet, for the low ability group, only one percentage point separates Conditions O and V. Performance under Condition S-V is even slightly worse than performance on the original passages. Yet, according to the Kincaid-Flesch formula (Kincaid et al., 1975), the passages had an average RGL of 5.5 compared to 11.5 for the original. This is the condition (i.e., the joint manipulation of vocabulary and sentence) that Klare (1976) predicted would have the maximum effect of on comprehension.

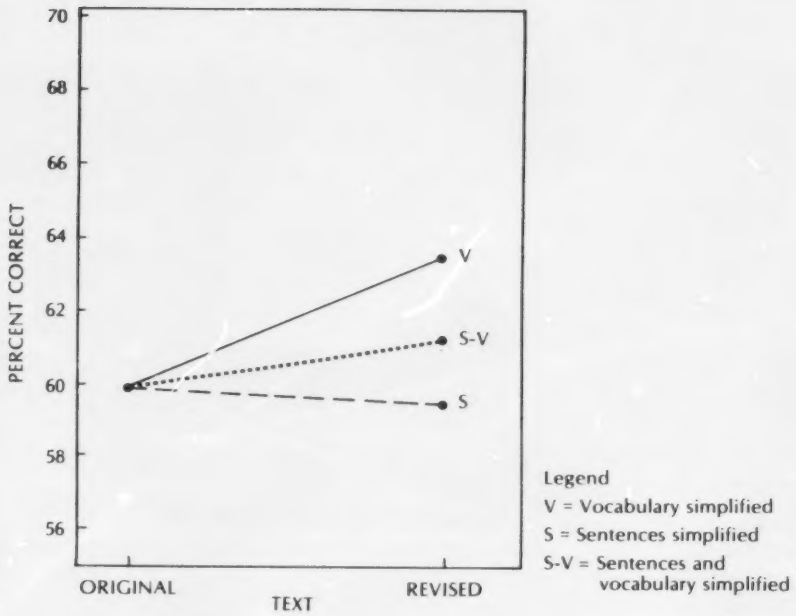


Figure 1. Percent correct under each revision condition, Experiment 1.

The results of this experiment offer no support for the hypothesis that readability formulas can be used as guides for simplifying text. Simplifying vocabulary and sentences has little if any effect on performance even though the readability, according to formula, is greatly improved. Even the small effects obtained are contrary to the predictions arising from the readability scores in Table 1. That is, while simplifying complex sentences produced the greatest improvement in readability score, the vocabulary simplification is what produced the positive, though trivial, performance effects. Even this effect is lost when the sentences are also simplified.

## EXPERIMENT 2

In Experiment 1, the passages were presented to allow for a reading rate of 100 words per minute. It was hypothesized that the absence of a speed component might be the factor accounting for the lack of a strong effect of increased readability on comprehension. Basically, given enough time most readers can struggle through a passage regardless of how difficult it is. Indeed, Klare (1976) reports that leisurely reading time is a primary factor in negating readability effects.

In this experiment, the reading-to-do task approach (i.e., a minimal memory requirement) was maintained but the speed factor was increased. This was accomplished by presenting the passages under the standard instructions of the Nelson-Denny reading test, which, like most reading tests, follows a reading-to-do approach. Questions and passages are presented together and the reader must find the answers to the questions. The Nelson-Denny has the additional characteristic of being a highly speeded test, allowing just over half of the test time (35 minutes) used in Experiment 1. Thus, following the test procedures provides a speeded reading-to-do task that should maximize the effects of the readability manipulations. Indeed, all factors identified in Klare's (1976) model are fixed to produce maximum readability effects.

## Method

A total of 571 recruits were tested, with 143 to 146 randomly assigned to each condition. After pretesting, participants received one of the experimental sets of booklets. In each booklet, a passage and its associated questions were on facing pages. Participants were given the published instructions for the Nelson-Denny test except that they were required to mark their answers on a machine-scorable answer sheet. Under the instructions, the participants are encouraged to move through the passages as quickly as possible, but they are free to skip passages or return to earlier passages.

## Results and Discussion

Recruits were divided into high, medium, and low ability groups based on the overall distribution of pretest scores. The low ability group consisted of those with scores below 10.0 (7.9 RGL); the high ability group, with scores above 16.0 (10.7 RGL); and the medium ability group, with scores between 10.0 and 16.0. Because the levels of the ability variable were defined in terms of the overall distribution, sample size across the 12 cells of the experiment was variable, ranging from 34 to 61.

Pretest scores were analyzed using a 3x2x2 unweighted means ANOVA. The main effect due to ability was highly significant ($F(2,559) = 270.6$, $p < .001$). The main effect of sentences was also significant ($F(1,559) = 6.6$, $p < .01$), indicating that the random assignment to conditions was not effective. An examination of the data indicates that performance in Conditions S and S-V was lower than that in Conditions O and V. Further, while the main effect is significant, the primary difference appears to be with the high ability group. The mean percent correct on the pretest for the sentence original and sentence simplified conditions was 37.2 and 37.5 percent respectively for medium ability participants, compared to 22.5 and 21.5 percent for low ability participants.

Interest was primarily on the medium and low ability readers since this is where the readability effects are expected. Since the bias in assignment seems to be primarily in the high ability groups, the pretest data were analyzed again using only the medium and low ability participants. The 2x2x2 unweighted means ANOVA yielded ability as the only significant effect ($F(1,376) = 850.8$, $p<.001$). The sentence effect did not even approach significance ($F < 1.0$).

Analysis of the experimental effects was carried out using only data from medium and low ability participants.[2] A 2x2x2 unweighted means ANOVA of the total number of items correct yielded significant main effects due to ability ($F(1,376) = 109.4$, $p < .001$) and vocabulary ($F(1,376) = 5.6$, $p < .02$). In addition, the vocabulary-by-ability interaction approached significance ($F(1,376) = 3.23$, $p< .08$). The interaction is in the right direction; the vocabulary effect is stronger with the low ability participants. However, the effect is minor, with the mean percent correct under conditions of vocabulary unchanged versus vocabulary simplified being 26.0 and 30.7 percent respectively for the low ability condition. This is, in fact, similar to the magnitude of the vocabulary effect in Experiment 1.[3]

---

[2] Analyses of covariance with ability as the covariate, which were carried on the entire data set in parallel to the ANOVAs, led to identical conclusions.

[3] An additional analysis carried out on the proportion of items correct to those attempted produced similar results.

The findings are displayed in Figure 2 with each revision condition contrasted to performance with the original passages. Again, the findings look very similar to those obtained in Experiment 1. Condition S yielded the poorest performance, though not significant, for both ability levels. Condition V facilitated performance only for low ability participants, but performance was increased by only five percentage points.

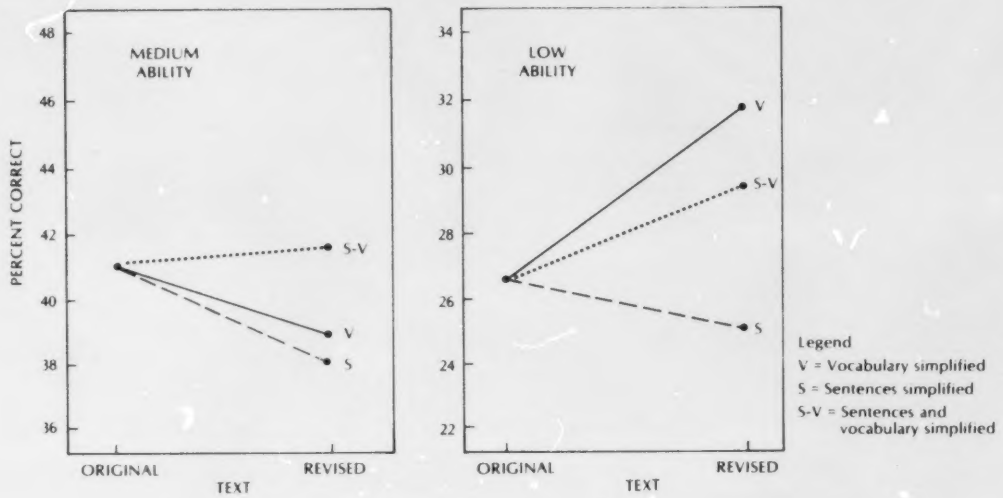


Figure 2. Percent correct under each revision condition for medium and low ability groups, Experiment 2.

Increasing the speed of the test should have increased the effects of vocabulary and sentence simplification (Klare, 1976). The test certainly was highly speeded; 50 percent of the participants did not respond to any of the last ten items and only 11 percent answered the last item. Yet, there is no practical effect of the more readable writing. An additional analysis was performed to determine if perhaps increased readability resulted in a lower error rate; that is, with the "difficult" test, readers could have given up and raced through the test primarily guessing. Under this strategy, they would increase the absolute number of items answered correctly but not the proportion correct. Thus, a 2x2x2 unweighted means ANOVA was carried out on the number correct relative to the number of items attempted. As with the analysis of the total number of correct items, the main effects of ability and vocabulary achieved significance ($F(1,376) = 28.9$, $p < .001$ and $F(1,376) = 8.7$, $p < .01$, respectively). None of the interactions achieved significance. As with the analysis of total correct, the vocabulary effect is of little practical significance, with only a difference of two percentage points between the means for the main effect (.58 for vocabulary simplified and .56 for vocabulary original).

As with Experiment 1, the present results offer no support to the hypothesis that the procedures of simplifying vocabulary and sentences by themselves will improve the comprehensibility of text materials. In this study, as with the previous one, the task and testing conditions were not especially motivating. In addition in this study, the test conditions were highly speeded. According to Klare (1976), the combination of low motivation, a highly speeded test, and low ability levels should maximize the likelihood that at least Condition S-V would produce a meaningful improvement in performance. For low ability participants, the materials were simplified from about three grade levels above their average reading skill to about two grade levels below their average reading skill. Yet, this large manipulation produced an effect that must be considered to be of no practical significance.

## EXPERIMENT 3

In Experiments 1 and 2, the test of comprehension was the ability to answer multiple-choice questions about the passages. It was hypothesized that the failure to find effects of practical significance may have been due to the use of the multiple-choice questions. The Nelson-Denny questions and procedures which are designed to distinguish between people with differing levels of reading skill, are apparently quite successful in doing so (Buros, 1968). However, recent research findings indicate that many questions on reading comprehension tests can be answered well above the chance level without the relevant passage being read (Tuinman, 1973-1974; Pyrczak, 1972). Thus, it would seem that many reading tests can successfully distinguish levels of reading skill with items that do not always assess reading skill. These items must be measuring individual difference variables that generally are highly correlated with reading skill.

Items that measure variables correlated with reading skill may be acceptable for assessing individual differences in skill. However, in Experiments 1 and 2, the intent was to measure differences in passages, not individuals. Here, questions that do not depend on reading the passage for a correct answer are clearly not acceptable. Comprehension differences due to difference in passage readability cannot be expected if the passages need not be read.

In this experiment, comprehension was tested by a cloze comprehension test (Taylor, 1953). The passages were presented to the participant with every fifth word deleted and the task was to fill in the deleted words. This is the procedure commonly used to assess comprehension in the contemporary development of readability formulas (Kincaid et al., 1975; Caylor, Sticht, Fox, & Ford, 1973). Once again, the experimental conditions were optimal for obtaining the predicted effects due to readability revisions (vocabulary and sentence simplification).

### Method

Only passages two and three of the Nelson-Denny test were used in this experiment. The mean Kincaid-Flesch (Kincaid et al., 1975) readability scores for these passages were 12.3, 10.9, 6.0, and 5.7 RGLs for Conditions O, V, S, and V-S respectively. Thus, the distribution of readability scores was about the same here as for the entire set of passages. The standard cloze procedure (Taylor, 1953), where every fifth word is deleted, was utilized. Each passage in each condition was prepared with five different cloze deletion patterns. That is, in one version, the 1st, 6th, 11th, etc. words were deleted, while in the next version, the 2nd, 7th, 12th, etc. words were deleted. This progression continued across all five versions until all words had been deleted.

The passages were presented on successive pages of a booklet. Participants were told to move to the second passage after completing the first. No time limit was imposed for completion, which is consistent with standard cloze testing procedures. In all, 124 recruits were tested, with 31 assigned to each condition.

## Results

Recruits were divided into high and low ability groups based on the overall median pretest score. The pretest performance was then analyzed with a 2x2x2 unweighted means ANOVA. Once again the results of the analysis indicated that the random assignment to groups did not result in equivalent groups within ability levels. In addition to the expected significance of ability ($F(1,166) = 229.0$, $p < .01$), the ability-by-sentence and the sentence-by-vocabulary interactions were also significant ($F(1,116) = 4.2$ and 7.9 respectively, $p<.05$). Examination of the means for the pretest shows that the significant interactions are due to group differences of up to seven percentage points at both ability levels. Thus, the attempt to assign participants randomly to readability conditions was not successful, and any subsequent ANOVA would have been uninterpretable. Instead, the experimental data were analyzed by an analysis of covariance with the pretest score serving as the covariate. The cloze performance was scored using both a stringent criterion and a lenient criterion. Under the stringent criterion, the blank had to be filled in with the exact word that had been deleted. Under the lenient criterion, synonyms were allowed. Two scorers were used and both had to agree that a response was a synonym. The cloze scores averaged across passages served as the basic data for analysis.

The analysis of covariance of exact cloze performance failed to yield any effects significant at the .05 level. However, the vocabulary-by-sentence interaction approached significance ($F(1,119) = 3.09$, $p<.10$). The analysis of covariance of synonym cloze scores yielded virtually identical results, with the vocabulary-by-sentence interaction once again approaching significance ($F(1,119) = 3.06$, $p<.10$). The adjusted means under the exact scoring criterion are presented in Figure 3, where it can be seen that the interaction is due to a slight decrease in performance when only vocabulary or sentences are simplified, but a slight increment in performance when both are simplified. The difference between Conditions O and S-V is only 1.5 percentage points. Once again, minimal effects due to simplification were obtained. The mean cloze scores are generally at the level considered to reflect an adequate level of comprehension (i.e., 38% cloze) (Bormuth, 1967; Rankin & Culhane, 1969). Thus, neither a basement nor a ceiling effect limited the range of the means. It would seem then that, under most conditions, revising materials by simplifying vocabulary and sentences is not an effective way of improving comprehension, at least when the reader is not required to organize and store the text.

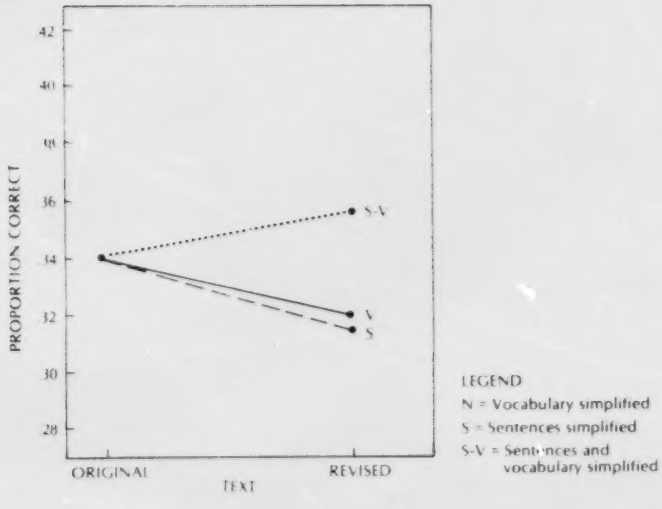


Figure 3. Proportion correct (adjusted means) under exact cloze scoring for each revision condition, Experiment 3.

## EXPERIMENT 4

In all three of the previous experiments, simplifying vocabulary and sentences failed to facilitate comprehension. This finding seems to contradict a long history of learning research in which it is a well established fact that frequently used and shorter sentences are "easier" than less frequently used words and longer sentences. However, the word frequency and sentence complexity factors are most strongly established in paradigms involving repeated learning trials and retention intervals.[4] Indeed, the research findings on word and sentence variables referenced in readable writing recommendations (e.g., Klare, 1975) come from studies of learning. It may be that these variables are relevant to memory, but not comprehension, or at least comprehension when memory is not required.

Without a memory requirement, there is only a minimal comprehension requirement. As Carroll (1972, p .7) has suggested: ". . . there is little use in comprehending a message unless the outcome of that comprehension is remembered and transfered to long term memory." Perhaps this statement should be tempered to suggest that a memory requirement will typically necessitate greater comprehension in the sense of deeper or more elaborate processing of the information.

Fass and Schumacher (1978) found that simplification of text facilitated comprehension when the readers did not know what information in the text had to be learned (i.e., memory was required). However, the simplification effects were reduced or eliminated by motivating or instructing subjects to engage in deeper processing activities. Thus, it may be that simplification in some way compensates for, or facilitates, the processing activities when required. In the first three experiments, however, there was no demand to organize, integrate, or store the information in the passage. In the cloze test, isolated phrases are studied. In the multiple-choice testing, the readers can skim the passage superficially looking for signals (e.g., key words) to direct them to the answer to the question. Thus, simplification may not have been effective simply because there were minimal processing demands.

In this experiment, the same materials were used and a memory component was added to the task in an attempt to increase the requirement for processing activities. Basically, in this experiment, a reading-to-learn task (Sticht et al., 1977) similar to the situation faced in the classroom was employed. Each passage was read with the participant not knowing the comprehension questions. Thus, "comprehending" the passage now meant acquiring all of the information presented.

### Method

A total of 244 recruits were tested, with 60 to 62 assigned to each of the four revision conditions. After pretesting, participants received one of the experimental booklets. The sequencing of materials in the booklets, the presentation procedures, and the time allowed for passage and questions were identical to those in Experiment 1, except that the questions on a passage did not precede the passage.

---

[4]Of course, variables like word frequency are related to the speed of word recognition, but this effect may not be relevant in the context of text reading. First, the effect is measured in milliseconds and hence may not have a practical effect. Second, the effect seems to be dependent on the absence of any context for the words to be identified (Broadbent, 1967).

## Results and Discussion

Paticipants were divided into high and low ability subgroups based on the median score of 15.3 (10.2 RGL) for the overall distribution of pretest scores. The pretest scores were subjected to a 2x2x2 unweighted means ANOVA to determine if the assumption of random assignment to readability conditions had been satisfied. The only factor to achieve significance in the analysis was the main effect due to ability ($F(1,236) = 421.3$, $p < .001$), which simply reflected the intended division of readers into ability groups.

The total number of correct items on the experimental passages was analyzed using a 2x2x2 unweighted means ANOVA. The main effects due to ability and vocabulary reached significance ($F(1,236) = 126.1$, $p<.001$ and $F(1,236) = 9.7$, $p<.002$, respectively). The interaction of vocabulary-by-ability also achieved significance ($F(1,236) = 6.3$, $p<.01$). The percent correct under each revision condition is shown separately in Figure 4 for high and low ability participants. The results for the high ability readers look virtually identical to the findings in the previous three experiments; there is virtually no effect when vocabulary is simplified and a slight decrease in performance when sentences are simplified.

The performance of low ability readers, in contrast to the previous findings, shows a strong effect due to vocabulary simplification--a 13-percent increase in performance. While not significant, the sentence manipulation also resulted for the first time in an improvement in performance. Thus, for low ability readers, simplification seems to be effective when a memory requirement is present. While the effects are significant, however, the findings are still not consistent with readability predictions. The vocabulary manipulation resulted in the smallest change in readability, but the largest effect on the comprehension test. The combination of vocabulary and sentence simplification is additive in its effects on readability formula scores (i.e., maximum readability improvement is obtained under Condition S-V). However, in each of the experiments, the effect on comprehension (or memory) performance seems to be an average of the individual effects (i.e., performance under Condition S-V falls between Conditions S and V). Thus, the conclusion must be that, even under the limited conditions where readability revisions result in improved performance, the formula score cannot be relied on to provide even a ranking of the relative difficulty of the revised material.

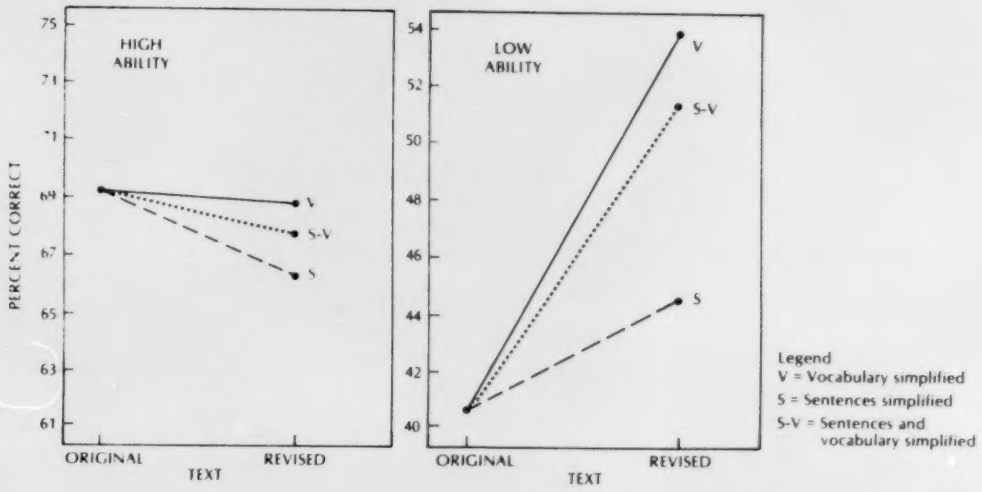


Figure 4. Percent correct under each revision condition for high and low ability groups, Experiment 4.

## EXPERIMENT 5

### Method

The purpose of this experiment was twofold. First, the correlational relationship of readability and comprehension was contrasted to the causative (or manipulated) effects of readability on comprehension. Second, the adequacy of the Nelson-Denny comprehension questions for reflecting readability effects was assessed. Results from the previous experiments suggest that revisions for readability are only effective when the comprehension test has a memory component. This, in turn, implies that the sentence and vocabulary variables are correlative but not causative variables in reading comprehension. That is, although simple sentences and simple vocabulary are characteristic of easy-to-comprehend material, they are not the factors underlying improved comprehension when no memory requirement is imposed. The relationship of readability and comprehension can be assessed by examining the correlation between the variables within experimental conditions in contrast to the causative effect that was assessed by comparing performance across conditions. The between-groups data fail to support a causative effect of sentence and word complexity on reading comprehension.

In this experiment, the same data were reexamined to determine if readability and comprehension covary within groups. This assessment served as a check on the adequacy of the Nelson-Denny test questions as measures of comprehension. As discussed in Experiment 3, the cloze testing has already supported the conclusions of Experiments 1 and 2. However, since the support is for the null hypothesis, there is still no assurance that the items were adequate for reflecting readability differences. An effect due to vocabulary simplification was obtained in Experiment 4. However, to the extent that the multiple-choice questions were passage independent (could be correctly answered without reading the passage), the potential magnitude of this effect, as well as a potential sentence effect, was suppressed. Thus, even in this experiment, the adequacy of the test items is not assured. However, if the questions are inadequate for reflecting between-group differences (i.e., they are passage independent), then they will also be inadequate for reflecting within-condition differences in readability and no correlation between the readability and comprehension scores for the eight passages would be expected.

The data from Experiments 1 and 4, the two experiments in which all participants answered all questions, were used for these analyses. Readability scores on each passage, calculated using the Kincaid-Flesch formula (Kincaid et al., 1975) served as one of the variables. The mean number of correctly-answered items on a passage in a particular readability condition served as the second variable. Since there were eight questions on the first passage, the mean score on this passage was halved for the analysis.

### Results and Discussion

The Kincaid-Flesch readability score for each of the eight passages under each revision condition is shown in Table 2. Each set of eight readability scores was used in the calculation of two Pearson product moment correlation coefficients, one using, as the second variable, the appropriate participants from Experiment 1; and the other, the participants from Experiment 4.

The resulting correlation coefficients shown in Table 3 were quite high except under Condition V--where only vocabulary was simplified. Thus, the data strongly support the adequacy of the Nelson-Denny test questions for reflecting differences in the readability of the passages. More correct answers were given on passages with easier readability scores.

Table 2

Kincaid-Flesch Readability Scores for Each Nelson-Denny Passage After Each Revision

| Passage | Original (O) | Revision Condition: Vocabulary Simplified (V) | Revision Condition: Sentences Simplified (S) | Revision Condition: Sentences & Vocabulary Simplified (S-V) |
|---|---|---|---|---|
| 1 | 8.5 | 7.2 | 4.2 | 3.4 |
| 2 | 9.4 | 8.2 | 6.1 | 4.5 |
| 3 | 12.3 | 11.2 | 5.5 | 3.8 |
| 4 | 13.3 | 15.4 | 11.2 | 8.2 |
| 5 | 8.2 | 6.7 | 6.3 | 5.2 |
| 6 | 13.2 | 9.4 | 8.1 | 6.5 |
| 7 | 10.4 | 7.6 | 8.7 | 5.8 |
| 8 | 16.6 | 14.8 | 7.9 | 6.6 |

Table 3

Correlation Between Passage Readability and Comprehension For Each Revision Condition in Experiments 1 and 4

| Revision Condition | Experiment 1 | Experiment 4 |
|---|---|---|
| Original (O) | -.70* | -.57 |
| Vocabulary simplified (V) | -.20 | -.07 |
| Sentences simplified (S) | -.65* | -.68* |
| Sentences and vocabulary simplified (S-V) | -.77* | -.84* |

Note. N = 8 (the 8 passages) for each Pearson product moment correlation.

*p < .05.

More importantly, these data add substance to the hypothesis that word and sentence difficulty are correlative, but not causative, factors in comprehension. The readability formula that basically measures word and sentence difficulty was predictive of comprehension in three of four conditions in two experiments. Yet, the manipulation of difficulty did not improve comprehension--except for vocabulary simplification when memory is required of low ability readers.

The findings from this study are not consistent with a similar analysis carried out by Entin and Klare (1978) using the passages from an alternative form of the Nelson-Denny test. They found a correlation coefficient of only .10 between performance on multiple-choice test items and passage readability. When the scores were corrected for passage independence, the correlation value increased to .41. Perhaps Entin and Klare obtained such low values because they allowed extensive reading time, even longer than in these experiments, and their readers were university students (i.e., high ability readers). Entin and Klare's second analysis, correcting for passage independence, suggests that a similar procedure carried out on the present data would result in even higher correlation coefficients between readability and comprehension. Yet, the results from Experiment 3, in which a cloze test was used, suggest that we still would not obtain effects due to the manipulation of readability.

## DISCUSSION AND CONCLUSIONS

In this series of experiments, the readability of passages was manipulated over wide ranges, from an RGL of 11.0 to 5.5, by following the most basic and fundamental rules for readable writing. The syntactic structure of sentences was simplified and more familiar words were substituted. Yet, except for one instance, these manipulations had no practical effect on comprehension, regardless of the reading skill of the participants. The experimental conditions met every requirement of Klare's (1976) model for producing comprehension effects. In one experiment (Experiment 3), the conditions of the development of the military readability formulas (Kincaid et al., 1975; Caylor et al., 1973), in the main, were replicated. Yet, even under these circumstances, readability manipulation did not affect comprehension.

Inadequate test questions cannot account for the findings since two test types (multiple-choice and cloze) failed to yield differences. The most suspect test type--multiple-choice--was the type with which significant comprehension effects were eventually obtained. In addition, performance on the multiple-choice items was strongly related to naturally occurring differences in readability. Thus, the questions were adequate for measuring readability effects.

The only instance in which the manipulations in readability had a practical effect on comprehension was when the comprehension quesions were not known at the time of reading (i.e., in the reading-to-learn format). Even here, the order of effect on comprehenson was not consistent with the order of readability differences. The smallest change in readability resulted in the greatest comprehension effect. However, the finding that simplification facilitates comprehension only on reading-to-learn tasks is consistent with the findings of Siegel et al. (1974). In that experiment, both vocabulary and sentences in technical materials were simplified. The materials were then used either as a correspondence training text, with a test after studying, or as a reference manual, with questions asked while the manual was available. It was only in the training situation--and then only when the manual was not supplemented by other instruction--that the effects due to simplification were obtained. The fact that simplification facilitates performance in reading-to-learn but not in reading-to-do indicates that the effect is not due simply to

improved comprehension of the individual units. Simplifying vocabulary does not improve performance simply by improving word knowledge. Simplifying sentences does not improve performance simply by improving comprehension of single sentences. If the effects were on isolated words and sentences, then clearly performance on the cloze test and in finding the answers to multiple-choice questions without a memory requirement would have been facilitated.

The requirements in the two types of tasks must be compared to determine the basis of the simplication effect. In reading-to-learn, the reader does not know what information must be retained to demonstrate comprehension. Thus, the material must be read with the goal of remembering everything that the reader judges to be important. This will most likely far exceed immediate memory capacity. Therefore, the task demands that the information in the passages be organized, integrated, and stored even for an immediate memory task. In reading-to-do, integration and organization typically are not required. Rather, the relevance of information is judged as reading progresses and can be discarded if it is not relevant to any of the specific goals. Thus, when simplifying vocabulary and sentences does facilitate comprehension, it appears to be through the organization and integration of the text. Perhaps the simplification results in more semantic associates and smaller chunks of information and, therefore, allows for greater flexibility in organizing and integrating the information. This in turn may result in storing of more information or in more efficient retrieval for answering the questions.

Fass and Schumacher (1978) found that instructing subjects to engage in deeper processing activities reduces the effects of simplifying text in a reading-to-learn situation. These findings, which are consistent with the hypothesis of this research, further suggest that providing guidance on how to process the text or motivation to carry out deeper processing can assist the reader in overcoming (i.e., comprehending) a difficult text, as well as simplifying the text for the reader.

In sum, text simplification seems to be one means of meeting the requirements for text processing. Thus, simplification should facilitate comprehension to the extent that deeper processing (i.e., integration and organization of the text) is required. This would include any reading-to-learn and reading-to-do tasks in which the reader must draw an inference, summarize the information, or perhaps carry away in memory a large string of facts (e.g., a procedure). Conversely, the effects of, or need for, simplification can be reduced by reducing the organizational demands. This could be done by providing an organizational structure, by motivating or instructing subjects to engage in appropriate processing activities, or by ensuring that the reader is knowledgeable in the content area. Klare (1976), for example, reports that, if the reader has knowledge of the content area, the effects of simplifying will be negated. Basically, content knowledge provides a schema (Anderson, Spiro, & Montague, 1977) or organizational structure for the information. A new structure or integration is not required.

The practical implication of the present findings is that a readability formula is not an effective production criterion, even when the writer does not deliberately write to the formula. Simplifications based on readable writing guidelines will be ineffective for technical manuals, at least where finding factual information is the task. For training manuals used in classroom instruction or in studying for advancement exams and for technical manuals read for "understanding," readable writing revision may have some effect under limited circumstances, but the effect will not be of the magnitude predicted by the readability formula score. Not only will formulas overstate the magnitude of a change, but they will not even effectively rank the difficulty of texts rewritten according to readable writing rules.

Of course, revisions based on readable writing guidelines can be effective at the extreme levels of diffiuclty. If the reader has no knowledge of the meaning of a significant proportion of the vocabulary, the sentences are extremely complex, and the reading task is more than to "look up," then a readability formula score can be an effective criterion for improving comprehension. That is, if the writer does not write to the formula.

In addition to being ineffective, the use of a readability formula seems to have limited the consideration of comprehension variables. More than just sentence and word factors determine comprehensibility, especially in technical manuals. Yet, as discussed earlier, comprehensibility specifications are presently limited to readability. Although graphics and the coordination of graphics and text play integral roles in technical manuals, readability analyses totally miss these factors. Procedural listing vs. paragraph presentation of information, highlighting techniques, and the organization of information within a paragraph are all variables affecting the comprehensibility of text that are independent of readability.

How are all of the comprehensibility factors to be considered in the production of a manual? There are three alternatives: guidelines, regulations, and changing the production system. Guidelines would not appear to be an effective approach. Although many such guidelines are quite logical (e.g., place the text and the relevant graphic on the same or facing pages), they are violated constantly. While there are innumerable books and training courses providing guidance for technical writers, comprehensibility is a continuing problem. Thus, guidance alone has proven ineffective. The use of a readability formula as a criterion is an attempt at objective regulation.

An alternative to the standard readability formula is a more complex formula that would include graphics, highlighting, and other diverse comprehensibility factors. Related to this notion is the recent preparation of a military standard for comprehensible writing of technical manuals (Department of Defense, 1978). This standard attempts to translate all the relevant research on comprehension into concrete writing and design statements. For example, the number of graphics per page, the use of procedural statements, and the use of specific highlighting techniques are so explicitly described that this document could be used as a criterion or specification for writing and design. While a specification of this complexity might be effective in increasing comprehensibility, it probably would not be cost effective by itself. Writers and designers would need training to use the specification and all details of the draft of the technical manual would have to be reviewed in relation to the explicit specifications. However, through a gradual evolution, including developing of training courses and programming the specifications into computer editing systems, a cost effective procedure for controlling the comprehensibility of manuals could be developed around such a standard.

A similar but more flexible system for comprehensibility control is embodied in McDonald-Ross and Waller's (1976) concept of a transformer in the text production process. The transformer is a group or individual whose sole responsibility is to ensure that the text (or technical manual) is maximally usable for the intended audience. The transformer is competent in educational technology, editing, graphic design, and the subject matter area. The transformer, then, is responsible for ensuring that principles like those embodied in the comprehension standard (Department of Defense, 1978) are applied appropriately. There should be a transformer function, or desk, in the production team and, when a contract is involved, the monitoring team.

The procurement of military hardware systems has traditionally encountered similar problems: The design process does not adequately attend to the manning requirements (i.e., the needs of the user). The Navy has recently included a transformer type office in the hardware procurement system (Chief of Naval Operations, 1977). This office, whose acronym is HARDMAN, has the sole function of reviewing each phase of the procurement effort to ensure that the "people considerations" are fully addressed. It is only through the institution of a complex specification or the institution of a transformer office analogous to the HARDMAN office that all aspects of a manual relevant to comprehension can be controlled.

## RECOMMENDATION

The text production process should be studied to determine how both government and contractor management of the production effort can be modified to increase attention to comprehensibility. This should include a consideration of computer-aided authoring systems, as well as changes in management priorities and in personnel qualifications. This study provided no data that would justify the use of a readability formula as a means of controlling the comprehensibility of text.

## REFERENCES


Adams, J. A. Human memory. New York: McGraw Hill, 1967.

Aiken, E. G. The mental quality of military personnel: Facts, theories, and remedies. Paper presented at the 16th Annual Symposium on Applied Military Psychology, The Netherlands, May 1980.

Anderson, R. C., Spiro, R. J. & Montague, W. E. Conference proceedings: Schooling and the acquisition of knowledge (NPRDC Tech. Rep. 78-6). San Diego: Navy Personnel Research and Development Center, December 1977 (AD-A049 878)

Battison, R., & Landesman, J. (Eds.) Simply Stated. Washington DC: Document Design Center, American Institute for Research, 1981.

Bormouth, J. R. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1967, 10, 291-299.

Broadbent, D. E. Word frequency effects and response bias. Psychological Review, 1967, 74, 1-15.

Buros, O. T. Reading: Tests and reviews. Highland Park, New Jersey: the Gryphon Press, 1968.

Carroll, J. B. Defining language comprehension: Some speculations. In J. B. Carroll and R. O. Freedle (Eds.) Language comprehension and the acquisition of knowledge. Washington DC: Winston and Sons, 1972.

Carver, R. P. Improving reading comprehension: Measuring readability. (Final Report, Contract No. N00014-72-C-0240.) Silver Springs, MD: American Institute for Research, 1974.

Caylor, J. S., Sticht, T. G., Fox, L. C., & Ford, J. P. Methodologies for determining reading requirements of military occupational specialties (HumRRO Tech. Rep. 73-5). Presidio of Monterey, CA: Human Resources Organization, March 1973.

Chief of Naval Operations, Military manpower vs hardware procurement (HARDMAN REPORT. Washington, DC: 1977.

Coleman, E. B. Improving comprehensibility by shortening sentences. Journal of Applied Psychology, 1962, 46, 131-134.

Dale, E. & Chall, J. S. A formula for predicting readability. Educational Research Bulletin, 21 January and 17 February 1948, 11-20; 37-54.

Department of the Air Force. Publication management writing understandable publications. HQ Operating Instruction 5.2. Washington, DC: Headquarters U.S. Air Force, March 1977.

Department of the Army. Improving the readability of Army publications. DA Circular 310-9. Washington, DC: Author, 15 December 1978.

Department of Defense. Military standard for technical manuals. DOD-STD-1685 (SH). Washington, DC: Author, 10 October 1978.

Department of Defense. Manuals, technical: General style and format requirements. MIL-M-38784A. Washington, DC: Author, December 1980.

Duffy, T. M., & Nugent, W. A. Reading skill levels in the Navy. (NPRDC Spec. Rep. 78-19). San Diego: Navy Personnel Research and Development Center, April 1978. (AD-A055 769)

Durell, D. D. Improved reading instruction. New York: Word Book Co., 1956.

Entin, E. B., & Klare, G. R. Some interrelationships of readability, cloze, and multiple choice scores on a reading comprehension test. Journal of Reading Behavior, 1978, 10, 417-436.

Fass, W., & Schumacher, G. M. Effects of motivation, subject activity, and readability on the retention of prose materials. Journal of Educational Psychology, 1978, 70, 803-807.

Flesch, R. F. A new readability yardstick. Journal of Applied Psychology, 1948, 32, 221-233.

Fry, E. B. A readability formula that saves time. Journal of Reading, 1968, 11, 513-516; 575-578.

Gunning, R. The technique of clear writing. New York: McGraw-Hill, 1952.

Goetz, E. T. Sentences in lists and in connected discourse (Tech. Rep. 3). Urbana-Champaign, IL: Laboratory for Cognitive Studies in Education, November 1975.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. Derivation of new readability formulas (automated readability index, fog count, and Flesch reading ease formula) for Navy enlisted personnel (Research Branch Report 8-75). Millington, TN: Naval Air Station, Memphis, February 1975.

Klare, G. R. Some suggestions for clear writing found in fifteen sourcebooks. Unpublished manuscript. Ohio University. Undated.

Klare, G. R. The measurement of readability. Ames, IA: Iowa State University Press, 1963.

Klare, G. R. A manual for readable writing. Glen Burnie, MD: REM Company, 1975.

Klare, G. R. A second look at the validity of readability formulas. Journal of Reading Behavior, 1976, 8, 129-152.

Klare, G. R. Readability standards for Army-wide publications (Evaluation Report 79-1). Fort Benjamin Harrison, IN: U.S. Army Administrative Center, 1979.

Kniffin, J. D., Stevenson, C. R., Klare, G. R., Entin, E., Slaughter, S., & Hooke, L. Operational consequences of a literacy gap (AFHRL-TR-79-22). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, November 1979.

McDonald-Ross, M. & Waller, R. The Transformer. The Penrose Annual, 1976, 69, 141-152.

Muller, W. Review of Navy research on the usability of technical manuals. In T. G. Sticht & D. Zapf (Eds.). Reading and readability research in the armed services (HumRRO FR-WD-CA-76-4). Alexandria, VA: Human Resources Research Organization, September 1976.

Nelson, M. J., & Denny, E. C. The Nelson-Denny reading test (Revised by J. I. Brown). Boston, MA: Houghton Mifflin Co., 1960.

Pressman, R. Legislative and regulatory progress on the readability of insurance policies. Washington, DC: American Institutes for Research. December 1979.

Pyrczak, F. Objective evaluation of the quality of multiple-choice test items designed to measure comprehension of reading passages. Reading Research Quarterly, 1972, 8, 62-71.

Rankin, E. F. & Culhane, J. W. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1969, 13, 193-198.

Siegel, A. I., Lautman, M., & Burkett, J. Reading grade level adjustment and auditory supplementation techniques for increasing textual comprehensibility. Journal of Educational Psychology, 1974, 66, 895-902.

Sticht, T. G., Fox, L. C., Hauke, R. N., & Zapf, D. W. The role of reading in the Navy (NPRDC Tech. Rep. 77-40). San Diego: Navy Personnel Research and Development Center, September 1977. (AD-A044 227)

Sulit, R. A., & Fuller, J. J. Navy technical manual system (NTMS) program summary (TM-186-76-1). Bethesda, MD: Naval Ship Research and Development Center, March 1976.

Taylor, W. L. Cloze procedures: A new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30,000 words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

Tuinman, J. J. Determining the passage dependency of comprehension questions in five major tests. Reading Research Quarterly, 1973-1974, 9, 206-223.

Tuinman, J. J., & Brady, M. E. How does vocabulary account for variance on reading comprehension tests? A preliminary to an instructional analysis. Presented at the National Reading Conference, Houston, TX: December 1973.

# APPENDIX

## ORIGINAL AND THREE REVISIONS OF ONE OF THE EXPERIMENTAL PASSAGES FROM THE NELSON-DENNY READING TEST

## ORIGINAL AND THREE REVISIONS OF ONE OF THE EXPERIMENTAL PASSAGES FROM THE NELSON-DENNY READING TEST

Original

The night was cloudy, and a drizzling rain, which fell without intermission, added to the obscurity. Steadily, and as noiselessly as possible, the Spaniards made their way along the main street, which had so lately resounded to the tumult of battle. All was now hushed in silence; they were only reminded of the past by the occasional presence of some solitary corpse, or a dark heap of the slain, which too plainly told where the strife had been hottest. As they passed along the lanes and alleys which opened into the great street, they easily fancied they discerned the shadowy forms of their foe lurking in ambush, ready to spring upon them. But it was only fancy; the city slept undisturbed even by the prolonged echoes of the tramp of the horses, and the hoarse rumbling of the artillery and baggage trains. At length, a lighter space beyond the dusky line of buildings showed the van of the army that it was emerging on the open causeway. They might well have congratulated themselves on having thus escaped the dangers of an assault in the city itself, and that a brief time would place them in comparative safety on the opposite shore.

Vocabulary Revised

The night was cloudy and a sprinkling rain, which fell without a break, added to the darkness. Without stopping, and with as little noise as possible, the Spaniards made their way along the main street, which had so recently roared to the noise of battle. All was now hushed in silence; they were only reminded of the past by the presence of a single dead body, or a dark heap of the dead, which too clearly told where the battle had been worst. As they passed along the lanes and alleys which opened into the great street, they easily fancied they saw the shadows of their enemy lying in wait, ready to spring upon them. But it was only fancy; the city slept without being bothered even by the constant sounds of the tramp of the horses, and the rough rumbling of the cannons and baggage trains. At length, a bright space beyond the dark line of buildings showed the look-out for the army that they were coming out on the open highway. They might well have rejoiced on having thus escaped the dangers of an attack in the city itself, and that a brief time would place them in greater safety on the opposite shore.

Sentences Revised

The night was cloudy. A drizzling rain added to the obscurity. It fell without intermission. The Spaniards made their way along the main street. They moved steadily and as noiselessly as possible. The street had so lately resounded to the tumult of battle. All was now hushed in silence. The occasional presence of some solitary corpse reminded them of the past. A dark heap of the slain was another reminder. Plainly, the strife had been hottest there. They passed along the lanes and alleys opening into the great street. They easily fancied the shadowy forms of their foe lurking in ambush. The enemy looked ready to spring upon them. But it was only fancy. The city slept undistrubed by the hoarse rumbling of artillery and baggage trains. Even the prolonged echoes of the tramp of horses did not disturb the city. At length, there was a lighter space beyond the dusky line of buildings. This informed the army van of their emergence onto the open causeway. They might well have congratulated themselves. They had thus escaped the dangers of an assault in the city itself. A brief time would place them in comparative safety on the opposite shore.

Sentences and Vocabulary Revised

The night was cloudy. A sprinkling rain added to the darkness. It fell without a break. The Spaniards made their way along the main street. They moved without stopping and with as little noise as possible. The street had so recently roared to the noise of battle. All was now hushed in silence. The presence of a single dead body reminded them of the past. A dark heap of the dead also reminded them. Clearly, the battle had been worst there. They passed along the lanes and alleys opening into the great street. They easily fancied the shadows of their enemy lying in wait. The enemy looked ready to spring upon them. But it was only fancy. The city slept without being bothered by the rough rumbling of the cannons and baggage trains. Even the constant sounds of the tramp of horses did not bother the city. At length, there was a bright space beyond the dark line of buildings. This informed the army look-out of their coming out onto the open highway. They might well have rejoiced. They had thus escaped the dangers of an attack in the city itself. A brief time would place them in greater safety on the opposite shore.

## DISTRIBUTION LIST

Military Assistant for Training and Personnel Technology (ODUSD(R&AT))
Director of Manpower Analysis (ODASN(M))
Chief of Naval Operations (OP-01), (OP-11), (OP-12) (2), (OP-13), (OP-14), (OP-15), (OP-115) (2), (OP-140F2), (OP-987H)
Chief of Naval Material (NMAT 0722), (NMAT 08L)
Deputy Chief of Naval Material (Technology)
Chief of Naval Research (Code 200), (Code 440) (3), (Code 442), (Code 448)
Chief of Information (OI-213)
Chief of Naval Education and Training (02), (022), (N-5)
Chief of Naval Technical Training (016)
Commander Fleet Training Group, Pearl Harbor
Commander Naval Air Systems Command
Commander Naval Electronic Systems Command
Commander Naval Facilities Engineering Command
Commander Naval Sea Systems Command
Commander Naval Military Personnel Command (NMPC-013C)
Commander Naval Air Development Center
Commander David W. Taylor Naval Ship Research and Development Center
Commanding Officer, Naval Education and Training Program Development Center (Technical Library) (2)
Commanding Officer, Naval Education and Training Support Center, Pacific
Commanding Officer, Naval Regional Medical Center, Portsmouth, VA (ATTN: Medical Library)
Commanding Officer, Naval Technical Training Center, Corry Station (Code 101B)
Commanding Officer, Naval Training Equipment Center (Technical Library)
Director, Career Information and Counseling School (Code 3W34)
Director, Defense Activity for Non-Traditional Education Support
Director, Management Information and Instructional Activity Branch Office, Memphis
Director, Naval Civilian Personnel Command
Director, Training Analysis and Evaluation Group (TAEG)
Officer in Charge, Central Test Site for Personnel and Training Evaluation Program
Officer in Charge, Naval Occupational Development and Analysis Center
Commander, Army Research Institute for the Behavioral and Social Sciences, Alexandria (PERI-ASL)
Commander, Headquarters TRADOC (ATPL-AE), Fort Monroe
Director, U.S. Army TRADOC Systems Analysis Activity, White Sands Missile Range (Library)
Chief, Army Research Institute Field Unit, Fort Harrison
Commander, Air Force Human Resources Laboratory, Brooks Air Force Base (Scientific and Technical Information Office)
Commander, Air Force Human Resources Laboratory, Lowry Air Force Base (Technical Training Branch)
Commander, Air Force Human Resources Laboratory, Williams Air Force Base (AFHRL/OT)
Commander, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base (AFHRL/LR)
Chief, Center Programs Division (AFMTC/XR), Lackland Air Force Base
Superintendent, U.S. Coast Guard Academy
Defense Technical Information Center (DDA) (12)